A CONTINGENCY MULTI-MICROPHONE NOISE REDUCTION STRATEGY BASED ON LINEARLY CONSTRAINED MULTI-CHANNEL WIENER FILTERING

Randall Ali, Marc Moonen

KU Leuven, Dept. of Electrical Engineering (ESAT)
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics
Kasteelpark Arenberg 10, 3001 Leuven, Belgium

ABSTRACT

The Minimum Variance Distortionless Response (MVDR) beamformer is a popular multi-microphone noise reduction and speech enhancement strategy that can be implemented either as a fixed-constraint MVDR beamformer, with a pre-defined Relative Transfer Function (RTF) or based on a Multi-channel Wiener Filter (MWF) estimate. However, each implementation is not fully robust within a dynamic acoustic environment. For instance, performance degradations exist for the fixed-constraint MVDR beamformer when the source is not in the constraint direction and also for the MWF when the estimated RTF is poor. In this paper, we propose a contingency noise reduction strategy that uses a Linearly Constrained MWF (LC-MWF) to combine the positive aspects of both implementations. We proceed to derive the LC-MWF in relation to the MVDR beamformer implementations and demonstrate through simulations that the LC-MWF is indeed an intermediary solution that encompasses a wider range of acoustic conditions.

Index Terms— Multi-Microphone Noise Reduction, Minimum Variance Distortionless Response, Multi-Channel Wiener Filter, Generalized Eigenvalue Decomposition.

1. INTRODUCTION

A popular multi-microphone noise reduction and speech enhancement strategy is the Minimum Variance Distortionless Response (MVDR) beamformer [1] [2]. In this technique, the response is preserved in a pre-defined constraint direction and is minimised in all other directions. The constraint direction can be the Acoustic Transfer Function (ATF) from the desired speech source location or more commonly in noise reduction applications, the Relative Transfer Function (RTF) [3]. While a fixed-constraint MVDR beamformer (MVDR-c) may prove to be an effective noise reduction strategy in some scenarios, its performance may degrade in other scenarios, particularly if there is a mismatch between the constraint direction and the actual speech source location [4].

A more recent strategy of growing interest is the Multi-Channel Wiener Filter (MWF) [5] [6], which is comprised of an MVDR beamformer followed by a single channel post-filter [7]. As opposed to using a fixed-constraint direction, this MVDR beamformer (MVDR-MWF) estimates the true RTF through the second order statistics of speech and noise signals (i.e. the corresponding correlation matrices). In some scenarios, this introduces an improvement in the performance of an MVDR-MWF over an MVDR-c, as it is not dependent on any a-priori information nor is it constrained to one particular direction. However, in low input Signal-to-Noise-Ratio (SNR) scenarios, when the correlated noise in the acoustic environment is larger than the speech signal, the ability to distinguish between periods of speech and non-speech activity becomes increasingly difficult. This can lead to a poor estimation of the true RTF, resulting in unreliable behaviour of an MVDR-MWF.

These two MVDR implementations rely on assumptions that are not always satisfied due to the variability of room acoustic conditions. Hence, with either implementation, it should be expected that there will be an ideal performance only for a subset of room acoustic conditions. For instance, an MVDR-c performs better in lower input SNR scenarios, where the speech source lies in the constraint direction and an MVDR-MWF performs better at higher input SNRs, regardless of the speech source location. Consequently, rather than using one of these implementations, we propose an integrated strategy that is a combination of both to arrive at an intermediary solution, facilitating a dynamic acoustic environment.

We accomplish this through the application of a Linearly Constrained MWF (LC-MWF), which introduces a linear constraint into the MWF cost function and a decision criterion for its implementation. With a neutralised single channel post-filter gain (unity), the LC-MWF simplifies into an MVDR-c always being active (constraint), and the inclusion of an MVDR-MWF only when there is a reliable estimation of the true RTF (decision). The LC-MWF can be perceived as a contingency noise reduction strategy. Whenever the acoustic environment results in a poorly estimated RTF (for instance, due to low input SNR, excessive reverberation time or non-stationary noises), we revert to an MVDR-c, which may perform better than an MVDR-MWF. When acoustic conditions are more favourable and the estimated RTF is more reliable, an intermediary performance between an MVDR-c and an MVDR-MWF will be achieved.

The data model and LC-MWF formulation are given in section 2. Section 3 elaborates on the proposed implementation through a Generalized Eigenvalue Decomposition (GEVD). Simulation results are provided in section 4 and conclusions in section 5.

2. DATA MODEL AND LC-MWF FORMULATION

For M microphones, we can represent the received signal in the frequency domain, $\mathbf{y}(\omega) = [\mathrm{y}_1\ \mathrm{y}_2\ \ldots \mathrm{y}_\mathrm{M}]^T$ as:

$$\mathbf{y}(\omega) = \mathbf{x}(\omega) + \mathbf{n}(\omega) = \mathbf{a}(\omega)S(\omega) + \mathbf{n}(\omega) \tag{1}$$

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), IWT O&O Project nr. 110722 'Signal processing and automatic fitting for next generation cochlear implants', IWT O&O Project nr. 150432 'Advances in Auditory Implants: Signal Processing and Clinical Aspects' and IWT O&O Project nr. 150611 'Proof-of-concept of a Rationed Architecture for Vehicle Entertainment and NVH Next-generation Acoustics (RAVENNA)'. The scientific responsibility is assumed by its authors.

where $\mathbf{x}(\omega)$ is the desired speech signal contribution, consisting of the ATF from the speech source location to the microphone array, $\mathbf{a}(\omega)$ and the speech signal, $S(\omega)$. $\mathbf{n}(\omega)$ represents the noise contribution. For brevity, we drop the function of frequency (ω) in the notation for the following derivations.

In the traditional MWF formulation, a set of filters (per frequency bin), $\mathbf{w} = [\mathrm{w}_1\ \mathrm{w}_2\ \ldots \mathrm{w}_\mathrm{M}]^T$ are designed to minimise the error between the filtered output signal and that of the unknown speech component in a chosen reference microphone. This can be extended to the speech distortion weighted MWF (SDW-MWF) [6], which incorporates a parameter, μ, that allows for a trade off between speech distortion and noise reduction. The corresponding cost function is given in (2), where $\mathbb{E}$ is the expectation operator, ^{H} is the Hermitian transpose and $\mathbf{e_{ref}}$ is a vector of length M, that contains a single entry in the first position (reference microphone) and the rest as zeros.

$$\min_{\mathbf{w}} \quad \mathbb{E}\{|\mathbf{w}^H\mathbf{x} - \mathbf{e}_{\mathbf{ref}}^H\mathbf{x}|^2\} + \mu\,\mathbb{E}\{|\mathbf{w}^H\mathbf{n}|^2\} \tag{2}$$

The filter that minimises (2) is then given by:

$$\mathbf{w}_{\text{sdw-mwf}} = (\mathbf{R_{xx}} + \mu\mathbf{R_{nn}})^{-1}\mathbf{R_{xx}}\,\mathbf{e_{ref}} \tag{3}$$

where the speech plus noise correlation matrix, $\mathbf{R_{yy}} = \mathbb{E}\{\mathbf{yy}^H\}$, the noise only correlation matrix, $\mathbf{R_{nn}} = \mathbb{E}\{\mathbf{nn}^H\}$ and the speech only correlation matrix, $\mathbf{R_{xx}} = \mathbb{E}\{\mathbf{xx}^H\} = \mathbf{R_{yy}} - \mathbf{R_{nn}}$, assuming that the speech and noise signals are uncorrelated. We also assume that $\mathbf{R_{xx}}$ is a rank-1 matrix consisting of a single speech source.

In our variation of the SDW-MWF, i.e. the LC-MWF, we introduce a linear constraint into (2):

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbb{E}\{|\mathbf{w}^H\mathbf{x} - \mathbf{e}_{\mathbf{ref}}^H\mathbf{x}|^2\} + \mu\,\mathbb{E}\{|\mathbf{w}^H\mathbf{n}|^2\} \\ \text{s.t.} \quad & \mathbf{w}^H\mathbf{d_c} = 1 \end{aligned} \tag{4}$$

where $\mathbf{d_c}$ is the RTF (with respect to the reference microphone) that defines the constraint direction for which the speech is to be preserved. In maintaining consistency with the formulation of the MWF, we can re-write this constraint as $\mathbf{w}^H\mathbf{d_c} = \mathbf{e}_{\mathbf{ref}}^H\mathbf{d_c}$, and subsequently re-write (4) as an unconstrained minimisation problem with an introduction of a weighting parameter, γ:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbb{E}\{|\mathbf{w}^H\mathbf{x} - \mathbf{e}_{\mathbf{ref}}^H\mathbf{x}|^2\} + \mu\,\mathbb{E}\{|\mathbf{w}^H\mathbf{n}|^2\} \\ & + \gamma\,\mathbb{E}\{|\mathbf{w}^H\mathbf{d_c} - \mathbf{e}_{\mathbf{ref}}^H\mathbf{d_c}|^2\} \end{aligned} \tag{5}$$

The filter that minimises (5) is then given by:

$$\mathbf{w}_{\text{lc-mwf}} = (\mathbf{R_{xx}} + \gamma\mathbf{d_c}\mathbf{d_c}^H + \mu\mathbf{R_{nn}})^{-1}(\mathbf{R_{xx}} + \gamma\mathbf{d_c}\mathbf{d_c}^H)\mathbf{e_{ref}} \tag{6}$$

The parameter $\gamma \in [0\ \infty]$, adjusts the weight to which the constraint is introduced. If $\gamma = 0$, then the SDW-MWF in (3) is obtained. If $\gamma \to \infty$ and $\mu = 0$, an MVDR-c is always active and the LC-MWF becomes a combination of an MVDR-c and an MVDR-MWF. In the following, we implement (6) through a GEVD that gives a non-trivial solution also when $\mu = 0$.

3. GEVD-BASED LC-MWF

The GEVD-based SDW-MWF, i.e. in computing (3), was introduced by Serizel *et al.* [8]. A GEVD-based computation of $\mathbf{R_{xx}}$ is used as a better alternative to simply subtracting $\mathbf{R_{nn}}$ from $\mathbf{R_{yy}}$, and this indeed leads to enhanced performance in low SNR scenarios.

The GEVD of the matrix pencil, $\{\mathbf{R_{yy}}, \mathbf{R_{nn}}\}$ (for an invertible $\mathbf{R_{nn}}$) is given by:

$$\mathbf{R_{nn}^{-1}R_{yy}} = \mathbf{U\Sigma_d U^{-1}} \tag{7}$$

where $\mathbf{\Sigma_d}$ is a diagonal matrix of the generalized eigenvalues (GEVLs), $\sigma_{\mathrm{d}1}, \sigma_{\mathrm{d}2} \ldots \sigma_{\mathrm{d}M}$, ordered such that $\sigma_{\mathrm{d}1} > \sigma_{\mathrm{d}2} > \ldots \sigma_{\mathrm{d}M}$, and $\mathbf{U}$ is the corresponding M x M matrix with the generalized eigenvectors (GEVCs), $\mathbf{u_1}, \mathbf{u_2} \ldots \mathbf{u_M}$ in the columns. This GEVD is also equivalent to a joint diagonalization of $\mathbf{R_{yy}}$ and $\mathbf{R_{nn}}$:

$$\mathbf{R_{yy}} = \mathbf{Q\Sigma_y Q}^H\ , \quad \mathbf{R_{nn}} = \mathbf{Q\Sigma_n Q}^H \tag{8}$$

where $\mathbf{Q}$ is a full-rank, M x M, invertible matrix, $\mathbf{\Sigma_y}$ = diag$\{\sigma_{\mathrm{y}1}, \sigma_{\mathrm{y}2} \ldots \sigma_{\mathrm{y}M}\}$ and $\mathbf{\Sigma_n}$ = diag$\{\sigma_{\mathrm{n}1}, \sigma_{\mathrm{n}2} \ldots \sigma_{\mathrm{n}M}\}$ are real valued diagonal matrices. Substituting (8) into (7), we deduce that $\mathbf{Q} = \mathbf{U}^{-H}$ and $\sigma_{\mathrm{d}j} = \frac{\sigma_{\mathrm{y}j}}{\sigma_{\mathrm{n}j}}$ (for $j = 1, 2 \ldots M$). The GEVD-based computation of $\mathbf{R_{xx}}$ then follows as:

$$\mathbf{R_{xx}} = \mathbf{Q}(\mathbf{\Sigma_y} - \mathbf{\Sigma_n})\mathbf{Q}^H = \mathbf{Q\Sigma_x Q}^H \tag{9}$$

In general, a rank N approximation to $\mathbf{R_{xx}}$ can be computed by selecting the first N GEVLs in the diagonal matrix $\mathbf{\Sigma_x} = \mathbf{\Sigma_y} - \mathbf{\Sigma_n}$. Hence, for a rank-1 $\mathbf{R_{xx}}$:

$$\mathbf{R_{xx}} = \mathbf{Q\Sigma_{xr1}Q}^H = \mathbf{q_1 q_1}^H\sigma_{\mathrm{x}1} \tag{10}$$

where $\mathbf{\Sigma_{xr1}}$ = diag$\{\sigma_{\mathrm{x}1}, 0, \ldots 0\}$, $\sigma_{\mathrm{x}1} = \sigma_{\mathrm{y}1} - \sigma_{\mathrm{n}1}$, corresponding to the largest GEVL and $\mathbf{q_1}$ is the corresponding first column of the matrix $\mathbf{Q}$. The filter as defined in (3) is then computed as:

$$\mathbf{w}_{\text{sdw-mwf}} = \mathbf{Q}^{-H}\left[\begin{array}{c|c} \frac{\frac{\sigma_{\mathrm{x}1}}{\sigma_{\mathrm{n}1}}}{\mu + \frac{\sigma_{\mathrm{x}1}}{\sigma_{\mathrm{n}1}}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]\mathbf{Q}^H\mathbf{e_{ref}} \tag{11}$$

Setting $\mu = 0$, results in the equivalent MVDR-type filter, with $\tilde{\mathbf{q}}_\mathbf{1}$ being the first column of $\mathbf{Q}^{-H}$:

$$\mathbf{w}_{\text{mvdr-mwf}} = \tilde{\mathbf{q}}_\mathbf{1}\mathbf{q_1}^H\mathbf{e_{ref}} \tag{12}$$

In order to extend this to a GEVD-based LC-MWF corresponding to (6), a new speech plus noise correlation matrix, $\mathbf{R_{yy+}}$ needs to be defined, such that:

$$\mathbf{R_{yy+}} = \mathbf{R_{yy}} + \gamma\,\mathbf{d_c d_c}^H \tag{13}$$

which implies that the new speech only correlation matrix is $\mathbf{R_{xx+}} = \mathbf{R_{xx}} + \gamma\,\mathbf{d_c d_c}^H$. The GEVD or joint diagonalization is then performed on the matrix pencil, $\{\mathbf{R_{yy+}}, \mathbf{R_{nn}}\}$, leading to diagonal matrix of GEVLs, $\mathbf{\Sigma_{d+}}$, with $\sigma_{\mathrm{d}1+} > \sigma_{\mathrm{d}2+} > \ldots \sigma_{\mathrm{d}M+}$, and $\sigma_{\mathrm{d}j+} = \frac{\sigma_{\mathrm{y}j+}}{\sigma_{\mathrm{n}j}}$ (for $j = 1, 2 \ldots M$). Formula (9) is then replaced by:

$$\mathbf{R_{xx+}} = \mathbf{Q_+}(\mathbf{\Sigma_{y+}} - \mathbf{\Sigma_n})\mathbf{Q_+}^H = \mathbf{Q_+\Sigma_{x+}Q_+}^H \tag{14}$$

However, here, the approximation to $\mathbf{R_{xx+}}$ is not necessarily rank-1. Firstly, large values of γ will determine the largest diagonal entry of $\mathbf{\Sigma_{x+}}, \sigma_{\mathrm{x}1+}$, and associated first column from $\mathbf{Q_+}, \mathbf{q_{1+}}$. In (13), if we set $\gamma >> \sigma_{\mathrm{y}1}$ (from (8)), then we can expect that $\sigma_{\mathrm{x}1+} \to \gamma$ and $\mathbf{q_{1+}} \to \mathbf{d_c}$, the constraint RTF. Hence a rank-1 approximation of (14) will (for large values of γ) result in an MVDR-c defined by $\mathbf{d_c}$. If the speech source lies in the constraint direction, this rank-1 approximation will be appropriate. On the other hand, if the speech source is not in the constraint direction, then some of this information will correspond to the second column of $\mathbf{Q_+}, \mathbf{q_{2+}}$. Hence, for such cases, a rank-2 approximation to (14) can be a better option:

$$\mathbf{w}_{\text{lc-mwf}} = \mathbf{Q}_{\mathbf{+}}^{-H} \left[\begin{array}{cc|c} \frac{\frac{\sigma_{\text{x1+}}}{\sigma_{\text{n1}}}}{\mu + \frac{\sigma_{\text{x1+}}}{\sigma_{\text{n1}}}} & 0 & \\ 0 & \frac{\frac{\sigma_{\text{x2+}}}{\sigma_{\text{n1}}}}{\mu + \frac{\sigma_{\text{x2+}}}{\sigma_{\text{n1}}}} & \mathbf{0} \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \mathbf{0} \end{array} \right] \mathbf{Q}_{\mathbf{+}}^{H} \mathbf{e}_{\mathbf{ref}} \quad (15)$$

However, if room acoustic conditions result in a poor estimation of the actual RTF, we may be better off with the contingency strategy of using only a rank-1 approximation. Consequently, the resulting LC-MWF may incorporate a decision parameter, $\alpha \in [0\ 1]$, as to whether or not to keep the rank-2 part of the approximation (with $\mu = 0$):

$$\mathbf{w}_{\text{mvdr-lc-mwf}} = (\tilde{\mathbf{q}}_{\mathbf{1+}}\mathbf{q}_{\mathbf{1+}}^{H} + \alpha\, \tilde{\mathbf{q}}_{\mathbf{2+}}\mathbf{q}_{\mathbf{2+}}^{H})\, \mathbf{e}_{\mathbf{ref}} \quad (16)$$

with $\tilde{\mathbf{q}}_{\mathbf{1+}}$ and $\tilde{\mathbf{q}}_{\mathbf{2+}}$ being the first and second columns of $\mathbf{Q}_{\mathbf{+}}^{-H}$ and $\mathbf{q}_{\mathbf{1+}}$ and $\mathbf{q}_{\mathbf{2+}}$ being the first and second columns of $\mathbf{Q}_{\mathbf{+}}$.

As for the decision parameter, $\alpha \in [0\ 1]$, our contingency strategy is as follows - the MVDR-c will always be active, but only when the speech source is not in the constraint direction and there is a reliable estimation of the true RTF do we incorporate the MVDR-MWF information (i.e. set $\alpha = 1$). One option for which to map the desired behaviour of α would be to observe the second GEVL of $\mathbf{\Sigma}_{\mathbf{d+}}$, $\sigma_{\text{d2+}} = \frac{\sigma_{\text{y2+}}}{\sigma_{\text{n2}}} = \frac{\sigma_{\text{x2+}}+\sigma_{\text{n2}}}{\sigma_{\text{n2}}}$, which is a measure indicative of the input SNR. If the input SNR is very low and/or if the speech source is in the constraint direction, then $\sigma_{\text{x2+}} \to 0$ and hence $\sigma_{\text{d2+}} \to 1$, indicating that the rank-2 part of (16) should be rejected. Therefore, when $\sigma_{\text{d2+}} > 1$, it suggests that, $\alpha \to 1$. However, in reverberant conditions this may not necessarily hold true, particularly when the speech source is in the constraint direction. If the reflections are strong enough, then they can result in $\sigma_{\text{d2+}} > 1$, falsely indicating that there is a speech source outside of the constraint direction with a reliable RTF estimation.

Consequently, we propose to use the ratio of $\sigma_{\text{d2+}}$ to the second GEVL from (7), σ_{d2}, as a more robust indicator:

$$\sigma_r = \frac{\sigma_{\text{d2+}}}{\sigma_{\text{d2}}} = \frac{\sigma_{\text{x2+}} + \sigma_{\text{n2+}}}{\sigma_{\text{x2}} + \sigma_{\text{n2}}} \quad (17)$$

Now, even if there is a very reverberant environment and the source is in the constraint direction, we can expect that the energy in both second GEVLs will be similar, resulting in $\sigma_r \to 1$. Therefore, only when $\sigma_{\text{x2+}} > \sigma_{\text{x2}}$, and hence $\sigma_r > 1$ would we expect that there is a good estimation of the RTF of a speech source outside of the constraint direction, suggesting to set $\alpha \to 1$.

To allow for a smooth transition of values of α from 0 to 1, as a function of σ_r, we propose to use a logistic function [9]:

$$\alpha = f(\sigma_r) = \frac{1}{1 + e^{-k(\sigma_r - \sigma_m)}} \quad (18)$$

where k is the steepness of the logistic curve and σ_m is transition point at which the logistic function has a value of 0.5. σ_m can be interpreted as the threshold for accepting or rejecting the information from the MVDR-MWF. In practice, these curves could be tuned to empirical data.

4. SIMULATIONS

The simulation environment consisted of a room with dimensions 7.1 m x 6.3 m x 5.2 m, a linear microphone array, a single speech source and ten localised noise sources. The array consisted of four omnidirectional microphones with an inter-element spacing of 4 cm. For the speech source signal, four sentences separated by silence from the English Hearing-In-Noise Test (HINT) database [10] were used. The localised noise source signals were uncorrelated excerpts of multitalker babble noise from Audiotec [11]. Uncorrelated white noise with a power of 5% of the speech signal power in the first microphone was also added to each of the microphone signals.

The localised noise sources were placed equidistantly from the centre of the microphone array, from angles 0^o to 180^o, at 20^o increments, with 0^o being the end-fire direction and 90^o as the broadside direction. The single speech source was placed 1 m away from the array, for different angles from 0^o to 180^o, at 20^o increments. For each position of the speech source, simulations to evaluate the MVDR-MWF, MVDR-c and LC-MWF were performed for different input SNRs and reverberation times (RTs). The input SNR was varied by changing the total gain of the localised noise sources.

The simulations were performed using the Weighted Overlap and Add (WOLA) method [12], with an FFT size of 256 and sampling frequency of 16 kHz. The room impulse responses were obtained with the image method [13] and implemented from [14]. A perfect voice activity detector (VAD) was also used to obtain the relevant correlation matrices. For the LC-MWF, the constraint direction ($\mathbf{d_c}$) was set to the end-fire direction, 0^o, $\gamma = 10^3$ (which was 10^3 times greater than the power of the maximum received input signal) and $k = 10$ and $\sigma_m = 1.5$ for the logistic function. The performance of the different noise reduction strategies was evaluated using speech intelligibility-weighted SNR improvement (Δ SI-SNR) and spectral distortion (SI-SD) measures defined in [6]. Figures 1 and 2 illustrate these performance metrics as a function of the source angle and input SNR for an RT = 0.3 s and an RT = 1 s respectively.

In figures 1(a) and 1(b), for an RT = 0.3 s, at the higher input SNR of 6 dB, when the speech source is in the constraint direction (0^o), the performance of the MVDR-MWF, the MVDR-c and the LC-MWF all converge. When the speech source is outside of the constraint direction, a compromise between the MVDR-c and the MVDR-MWF is achieved. In particular, the LC-MWF does not suffer from the excessive distortion as the MVDR-c. As the input SNR decreases (from figures 1(c) and 1(d) to 1(e) and 1(f)), the LC-MWF then gradually reverts to the MVDR-c (the contingency strategy). Now it is observed that the LC-MWF has a better performance over the MVDR-MWF in the constraint direction. In figure 2, for an RT = 1 s, a similar trend is noted to that of when the RT = 0.3 s, suggesting that σ_r can indeed be a robust indication into transitioning between the MVDR-c and MVDR-MWF.

5. CONCLUSIONS

We have developed a contingency multi-microphone noise reduction strategy that uses a Linearly Constrained Multi-Channel Wiener Filter (LC-MWF). When there is a reliable estimation of the true Relative Transfer Function (RTF) of the speech source, the LC-MWF is a combination of a fixed-constraint MVDR beamformer (MVDR-c) and an MVDR beamformer from an MWF estimate (MVDR-MWF). In adverse room acoustic conditions, when the RTF estimate is poor, the MVDR-MWF is rejected and the LC-MWF reverts to the MVDR-c (contingency strategy). The degree of reliability of the estimated RTF is based on a ratio of generalized eigenvalues (GEVLs) and is incorporated into a decision function for the LC-MWF. Preliminary simulation results have demonstrated that the LC-MWF has an intermediate performance between the MVDR-c and the MVDR-MWF, and is consistent for different reverberation environments. These simulations have excluded the potential voice activity detector

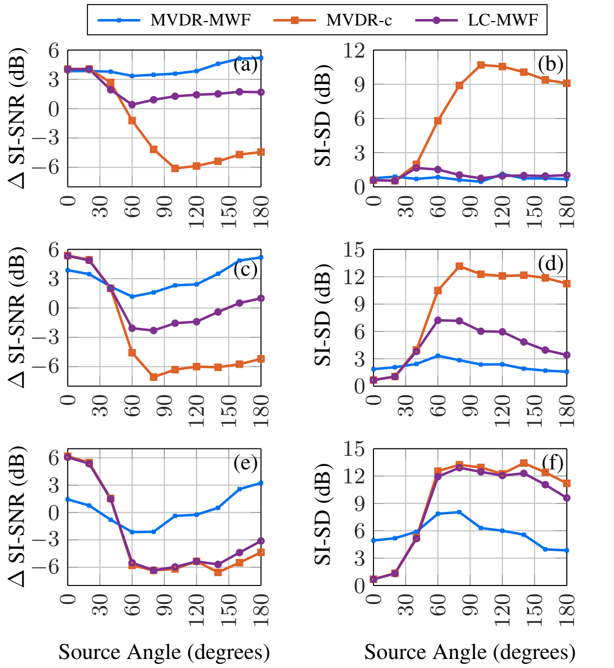

Fig. 1: MVDR-MWF, MVDR-c and LC-MWF performance. RT = 0.3 s. Graphs (a) and (b) are for an input SNR = 6dB, (c) and (d) for an input SNR = -6dB and (e) and (f) for an input SNR = -16dB.

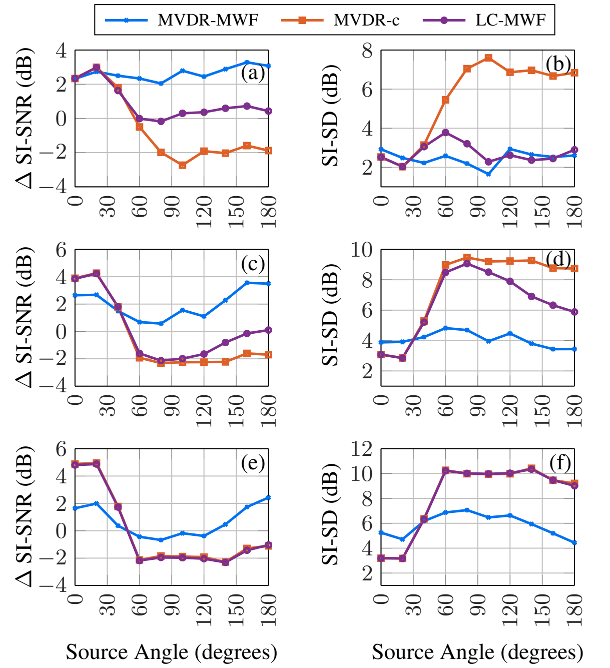

Fig. 2: MVDR-MWF, MVDR-c and LC-MWF performance. RT = 1 s. Graphs (a) and (b) are for an input SNR = 6dB, (c) and (d) for an input SNR = -6dB and (e) and (f) for an input SNR = -16dB.

(VAD) errors due to low input Signal to Noise Ratio (SNR). While such an approach has provided an initial and positive insight into the performance of the LC-MWF, future work will be geared towards a further analysis and evaluation of the strategy using an imperfect VAD. We also intend on reducing the computational complexity of the LC-MWF for practical implementations.

6. REFERENCES

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[2] E. Habets, J. Benesty, S. Gannot, and I. Cohen, *Speech Processing in Modern Communication: Challenges and Perspectives*, chapter 9: The MVDR Beamformer for Speech Enhancement, pp. 225–254, Springer, Berlin, Heidelberg, 2010.

[3] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech," vol. 49, no. 8, pp. 1614–1626, 2001.

[4] Chao Pan, Jingdong Chen, and Jacob Benesty, "Performance Study of the MVDR Beamformer as a Function of the Source Incidence Angle," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 1, pp. 67–79, 2014.

[5] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002.

[6] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.

[7] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, *Handbook on Array Processing and Sensor Networks*, chapter 10: Acoustic Beamforming for Hearing Aid Applications, pp. 269–302, John Wiley & Sons, Inc., 2010.

[8] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank Approximation Based Multichannel Wiener Filter Algorithms for Noise Reduction with Application in Cochlear Implants," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 785–799, 2014.

[9] Nicolas Bacaër, *A Short History of Mathematical Population Dynamics*, chapter Verhulst and the logistic equation (1838), pp. 35–39, Springer, London, 2011.

[10] M. Nilsson, S. D. Soli, and J. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise.," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994.

[11] Auditec, "Auditory Tests (Revised), Compact Disc, Auditec, St. Louis," St. Louis, 1997.

[12] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/Synthesis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.

[13] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943, 1979.

[14] E. Habets, "Room impulse response (RIR) generator,," 2010.